

Observer variation in the interpretation of Gram-stained urethral smears

Implications for the diagnosis of non-specific urethritis

J R WILLCOX, M W ADLER, AND E M BELSEY

From the Academic Department of Genitourinary Medicine, Middlesex Hospital Medical School, London

SUMMARY A study was carried out to determine whether the diagnosis of non-specific urethritis was affected by differences in the microscopical interpretation of urethral smears between individual observers (interobserver variation) and the same observer on separate occasions (intraobserver variation). A marked degree of both intraobserver and interobserver variation was found which—depending on the diagnostic criteria adopted—could affect both the diagnosis and treatment of many patients attending a clinic of genitourinary medicine.

Introduction

Although non-specific urethritis (NSU) is the most commonly notified sexually transmitted disease in England and Wales,¹ the criteria used to make the diagnosis vary from clinic to clinic.² The main individual criterion is based on positive microscopical findings, but again there is considerable variation between clinics in the number of leucocytes per high power field (hpf) accepted as being diagnostic of NSU.²

One factor that has received little consideration is variation in microscopical interpretation between individual observers (interobserver variation) and by the same observer on separate occasions (intraobserver variation). The present study was designed to examine how this might affect the diagnosis of NSU.

At this clinic the collection of urethral specimens, their staining, and microscopy is carried out by male nurses experienced in these procedures. The nurses report on the presence or absence of Gram-negative intracellular diplococci, polymorphonuclear leucocytes (PMNL), and other organisms. The slides are examined under oil immersion at $\times 100$ magnification ($\times 10$ eyepiece, $\times 100$ objective). The finding of no PMNL in at least three high power fields is reported as negative, of 1-9 as \pm , of 10-25 as $+$, and of >25 as $++$.

The physician will take into account the clinical history, findings on examination, and the results of the two-glass urine test and microscopy when establishing a diagnosis of NSU. In most instances, however, any patient with 10-25 PMNL/hpf ($+$) or >25 PMNL/hpf ($++$) on urethral smears will be treated for NSU and those with 1-9 PMNL/hpf (\pm) or no PMNL/hpf ($-$) will not.

Material and methods

During the study one of us (JRW) collected 60 Gram-stained male urethral slides which, in his opinion, consisted of 20 slides each showing \pm or $-$, $+$, or $++$ readings. In an attempt to provide variety and to serve as a guideline to the experience of the individual microscopist eight slides were included which showed typical Gram-negative intracellular diplococci. The 60 slides were then ordered randomly and numbered individually from 1 to 60.

Four male nurses experienced in microscopy kindly agreed to take part in the study. They were informed initially that they would be required to examine 60 slides each (20 at a time on different days). It was stressed to them that the study was not a test of their ability but purely an exercise to study individual variation. They were asked not to collaborate with their colleagues on their findings and to interpret the slides in their usual way. Each nurse used the same microscope.

After each nurse had read the 60 slides they were told that it was necessary for them to examine a further 60. Unbeknown to the nurses, each group of 60 contained the same slides, but the order was

Address for reprints: Dr J R Willcox, Special Treatment and Diagnostic Clinic, Plymouth General Hospital, Freedom Fields, Plymouth

Accepted for publication 29 August 1980

reversed in case a pattern of results was recognised which would perhaps lead to bias in reporting.

It was thought necessary to exclude the eight slides with intracellular Gram-negative diplococci from the analysis since most of them showed a ++ reading. An observer could be influenced in reporting, once typical Gram-negative diplococci had been seen. In the duplicated readings all four observers agreed that 10 of the 16 slides showed positive results. Two nurses both reported one slide as showing a negative result and each reported two further slides as showing negative results. One nurse reported one negative result and one had no false-negative results.

Thus interobserver and intraobserver variation was assessed from the results for each nurse reporting independently on the same 52 slides on two occasions.

STATISTICAL ANALYSIS

Intraobserver variation was analysed by calculating Spearman's rank correlation coefficient. A non-parametric multiple correlation coefficient, Kendall's coefficient of concordance, was used to assess interobserver variation. However, this statistic will take low values only when there is general disagreement between the observers under comparison, and could therefore mask the fact that one observer differed from the consensus view held by the rest. To examine this possibility non-parametric analyses of variance were also carried out.

Results

INTRAOBSERVER VARIATION

The intraobserver variation for the number of PMNL reported by the four nurses (observers 1, 2, 3, and 4) between the first and second groups of 52 slides is shown in table I. The calculated values of Spearman's rank correlation coefficient ranged from 0.64 to 0.76, and all were significantly greater than zero ($P < 0.0005$).

This variation can be seen in table I. For example, observer No 1 reported 35 out of 52 of the slides as

having the same numbers of PMNL on the two consecutive occasions; 10 slides were reported differently (for example, + on the first occasion and ++ on the second or vice versa), but this would not have affected treatment. More significantly, seven slides were reported as showing differences on the two occasions which would affect treatment (for example, + or ++ on the first and ± or - on the second occasion or vice versa). The corresponding figures affecting treatment for the other three observers (Nos 2, 3, and 4) were eight, nine, and 11 slides respectively.

INTEROBSERVER VARIATION

The interobserver variation between the four nurses is shown in table II. The level of agreement between them was significantly greater than zero ($P < 0.0001$) for both the first and second groups of 52 slides. When non-parametric analysis of variance was carried out, no significant differences were found between the four nurses in interpreting the second group of 52 slides. However, for the first group one nurse reported greater numbers of PMNL than the other three ($P < 0.01$).

TABLE II *Interobserver variation and effect on treatment*

Slides	With different result and effect on treatment					
	With same result		No effect		Some effect	
	No	%	No	%	No	%
1st group (52)	19	36.5	16	30.8	17	32.7
2nd group (52)	22	42.3	9	17.3	21	40.4

In the first group of 52 slides, 19 were reported identically by each of the four nurses, whereas 16 were reported differently (for example, + by one observer and ++ by the others), although treatment would not have been affected. A further 17 (32.7%) of the first group of slides were reported differently by one or more observers to a degree that would have affected treatment; the corresponding figure for the second group of slides was 21 (40.4%).

Discussion

Whatever the aetiology of NSU it is probable that microscopy of urethral smears will continue to be the cornerstone of its diagnosis. Some of the potential discrepancies in the interpretation of smears using different microscopes have been pointed out,³ but little work has been done to establish a standard technique for the taking and spreading of urethral smears, which can presumably affect the PMNL count whatever diagnostic criteria are used. It has

TABLE I *Intraobserver variation and effect on treatment (between two groups of 52 slides)*

Observer No	With different result and effect on treatment					
	With same result		No effect		Some effect	
	No	%	No	%	No	%
1	35	67.3	10	19.2	7	13.5
2	36	69.2	8	15.4	8	15.4
3	35	67.3	8	15.4	9	17.3
4	33	63.5	8	15.4	11	21.1

been shown that the time since urine was last voided can affect microscopical findings and hence the diagnosis of NSU,⁴ but no previous studies have considered the effects of intraobserver and inter-observer variation on PMNL reporting.

The results of this study have shown that even an experienced observer is surprisingly inconsistent in reporting on the same slide on two consecutive occasions, and the opinion of different observers looking at the same slide is even more variable.

Although in this study, no attempt has been made to equate the results with symptomatology and clinical follow-up, further work is planned in which this will be done. However, it is known that many patients with NSU have no symptoms.⁵ Thus it seems possible that in many patients it is purely chance that leads to the diagnosis of NSU, and the results of microscopy should perhaps be viewed with greater circumspection.

We are grateful to the nurses in the clinic for taking part in this study and allowing us to publish the findings. Particular thanks are due to Mr J Nash, the nursing officer, for his co-operation.

References

1. Department of Health and Social Security. *On the State of the Public Health. Report of the Chief Medical Officer for 1977*. London: HMSO 1978;59.
2. Adler MW. Diagnostic, treatment and reporting criteria for non-specific genital infection in sexually transmitted disease clinics in England and Wales. I Diagnosis. *Br J Vener Dis* 1978; **54**:422-7.
3. Aumonier FJ. Some problems in microscopy encountered in clinics for sexually transmitted diseases. *Br J Vener Dis* 1978; **54**:346-7.
4. Simmons PD. Evaluation of the early morning smear investigation. *Br J Vener Dis* 1978; **54**:128-9.
5. Rodin P. Asymptomatic non-specific urethritis. *Br J Vener Dis* 1971; **47**:452-3.